The Brain as a Hierarchical Organization

Isabelle Brocas
USC and CEPR

Juan D. Carrillo
USC and CEPR

Preliminary: May 2005

What is "Neuroeconomic Theory"?

Stylized evidence of Neuroscience:
which brain system is activated when

+

Modeling techniques of Micro theory:
agency & incentive theory,
organizational design, etc.

→ The brain is and should be modeled as
a multi-agent organization

Objective of this research

Understand behaviors difficult to reconcile with traditional theories (just as recent behavioral economics literature):

- Guilt
- Mistaken consumption of habit-forming goods, etc.

Provide "micro-microfoundations" for characteristics traditionally considered exogenous:

- Discounting
- Risk-aversion, etc.

Revisit the individual decision-making paradigm
(not decision-theory but game-theory approach)

This paper

Incorporate in a model of the brain two findings that have received support in neuro-experiments :

1. Conflict in the brain between [McClure et al. (2004)]
   - Forward-looking system (pre-frontal cortex) capable of intertemporal tradeoffs
   - Myopic system (paralimbic cortex) interested only in immediate gratification
2. Restricted cognitive access within brain to:
   - Motivations
   - Beliefs

"The heart has its reasons which reason knows nothing of"
(Blaise Pascal)

A caveat

1. Assumptions based on neuroscience evidence:
   - Conflict between myopic and forward-looking
   - Asymmetric information

2. Modeling choices (no evidence yet)
   - Vertical hierarchy:
     - Forward-looking = planner
     - Myopic = doer
   - Private information possessed by myopic

Related literature

1. Hyperbolic discounting with incomplete information
   (Carrillo-Mariotti, Brocas-Carrillo, Benabou-Tirole, Amador-Werning-Angeletos)

   Main Differences:

   - Conflict **within** (rather than between) periods
   - Asym. info **within** (rather than between) periods

2. Other dual-self theories
   (Thaler-Shefrin, Fudenberg-Levine, Loewenstein-O'Donoghue, Benhabib-Bisin, Bernheim-Rangel)

   Main Differences:

   - **Asym. Info** (rather than full info.) within periods
   - **Constraints** (rather than costs) in decision-making

The model

- 2 periods of consumption and labor (c_1, n_1) and (c_2, n_2)
- Utility

"Principal" **P**
prefrontal cortex

$$[\theta_1 u(c_1) - n_1] + [\theta_2 u(c_2) - n_2]$$

"Agent 1" $\mathbf{A_1}$
paralimbic cortex at date 1

$$\theta_1 u(c_1) - n_1$$

"Agent 2" $\mathbf{A_2}$
paralimbic cortex at date 2

$$\theta_2 u(c_2) - n_2$$

where $u' > 0$, $u'' < 0$ and θ_t is valuation at date t

- Consumption is non-negative: $c_t \geq 0$
- Labor is non-negative and bounded: $n_t \in [0, \bar{n}]$
- 1 unit of labor → 1 unit of income → 1 unit of consumption
- Perfect capital markets with interest rate $r > 0$
  ⇨ Intertemporal budget constraint:

$$c_1(1+r) + c_2 \leq n_1(1+r) + n_2$$

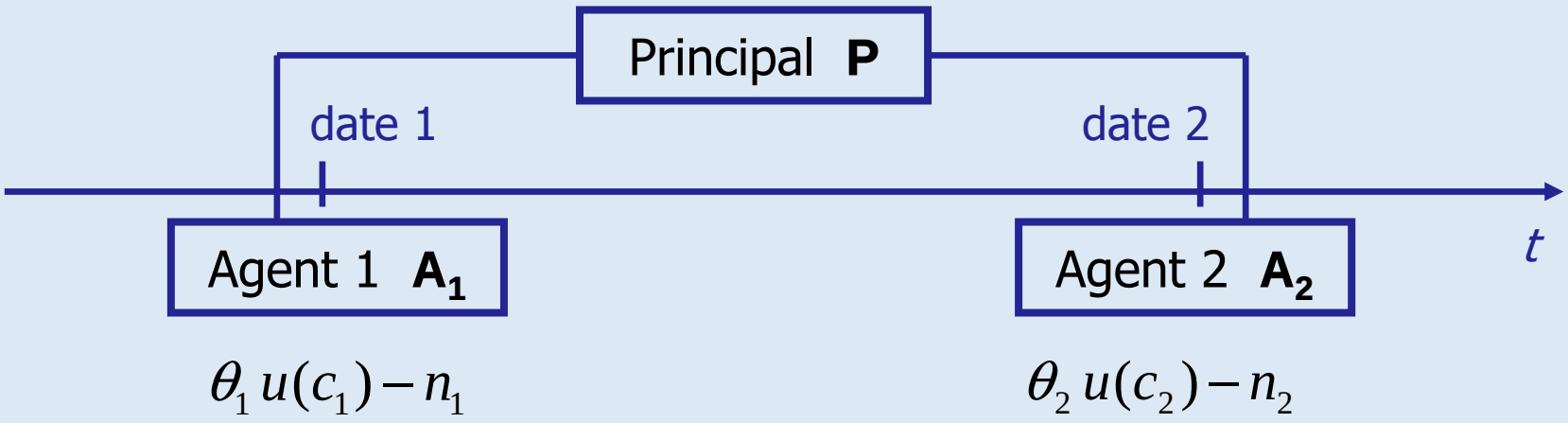
[Note: no individual rationality constraint]

- $\mathbf{A}_t$ chooses his preferred pair (c_t, n_t)

… but **P** can restrain $\mathbf{A}_t$'s choices
and we allow any conceivable rule / restriction

- **P** deals with $\mathbf{A}_1$ and $\mathbf{A}_2$ sequentially

$$[\theta_1 u(c_1) - n_1] + [\theta_2 u(c_2) - n_2]$$



Benchmark: conflict under full information

P knows the valuation θ_t of $\mathbf{A}_t$

For each θ_t, **P** imposes on $\mathbf{A}_t$ a specific pair $(c_t^o(\theta_t), n_t^o(\theta_t))$

Main characteristics of $(c_t^o(\theta_t), n_t^o(\theta_t))$:

- Consumption at t increases with θ_t (valuation at t)
- Labor at 1 is maximum (positive interest rate on savings)
- Labor at 2 is adjusted to meet budget constraint

⇨ Positive relation consumption at 1+2 and labor at 1+2
"work more in your lifetime to consume more in your lifetime"

⇨ No relation consumption at 1 and labor at 1

Conflict under asymmetric information

A$_t$ knows his valuation θ_t

P only knows that θ_t i.i.d. $F(\theta_t)$

⇨ **P** cannot impose restrictions that depend on valuation θ_t

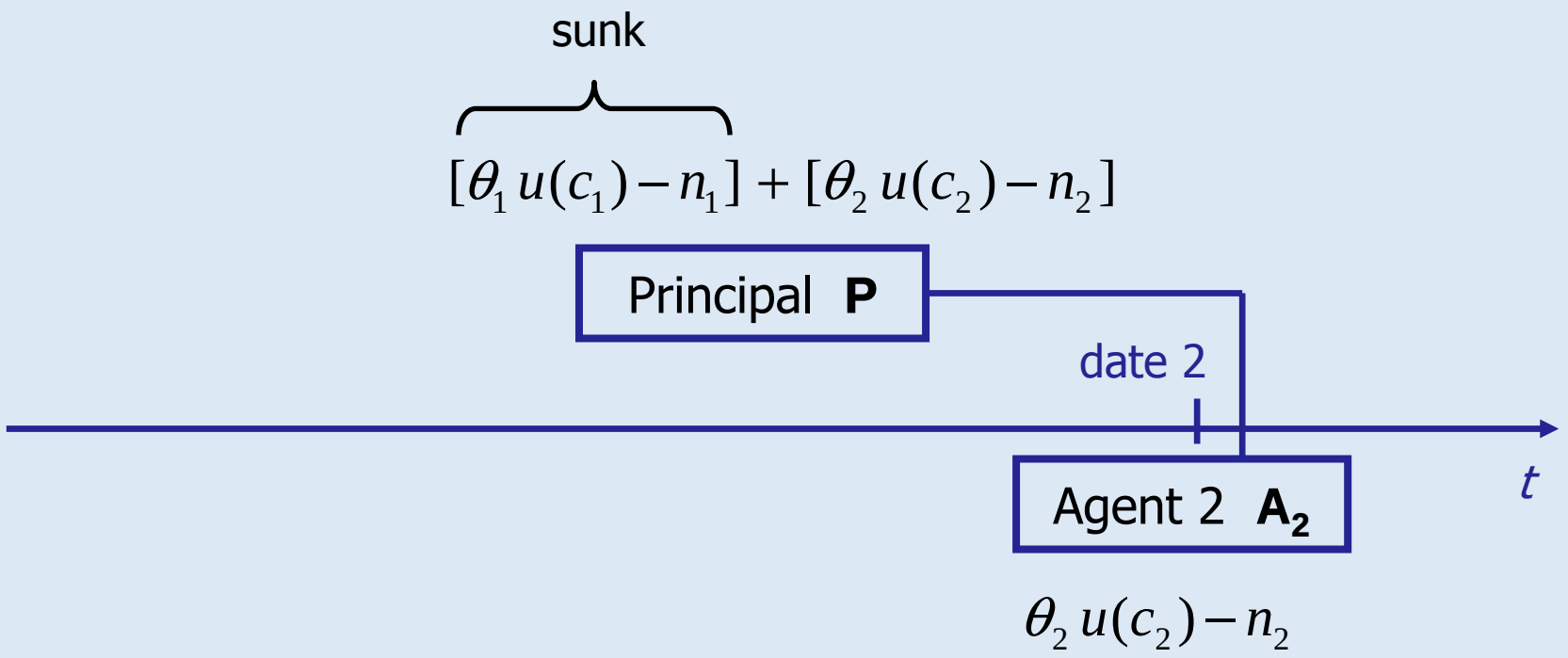
Note:

Because constraint (no access to θ_t) instead of cost of imposing choices:

- No presupposed tradeoff
- No preconceived idea of which restriction **P** will impose

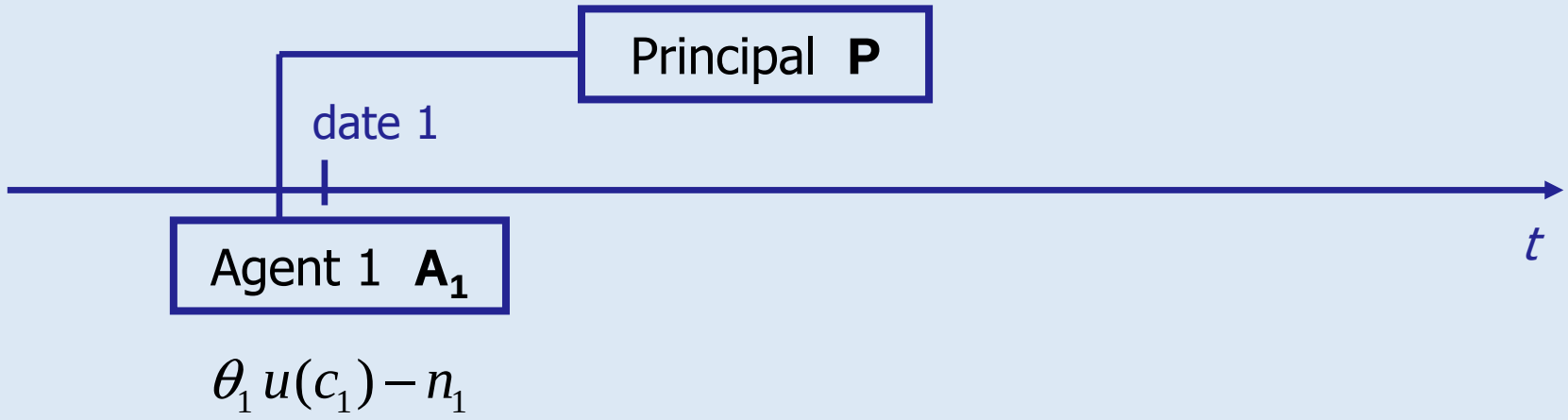
Optimal rule at date 2: P vs. A_2

Trivial. No restrictions (except budget balance) because at date 2 no conflict between **P** and $\mathbf{A}_2$



Optimal rule at date 1: P vs. A_1

$$[\theta_1 u(c_1) - n_1] + [\theta_2 u(c_2^*(\theta_2)) - n_2^*(c_1, n_1, c_2^*(\theta_2))]$$



$$\theta_1 u(c_1) - n_1$$

- **P** offers a menu of consumption and labor pairs

$$(c_1^*(\theta), n_1^*(\theta)),\ (c_1^*(\theta'), n_1^*(\theta')),\ (c_1^*(\theta''), n_1^*(\theta''))$$

  with as many pairs as possible valuations
- $\mathbf{A}_1$ picks the pair he prefers
- The pairs are designed such that
  - Different valuations ⇨ different choices
  - Higher valuation ⇨ more consumption and more labor

$$\theta > \theta' \quad \Rightarrow \quad c_1^*(\theta) > c_1^*(\theta') \quad \text{and} \quad n_1^*(\theta) > n_1^*(\theta')$$

  [the result is reminiscent of mechanism design literature]

Qualitative conclusions

- Endogenous emergence of (second-best) self-imposed rule: "work more today if you want to consume more today"
- Behavior looks like "guilt" without assuming it
- Current reward (leisure) tracks current earning (one day at a time effect)
- No consumption smoothing.
  Distribution of consumption over life cycle depends on:
  - Source of income (endowment vs. current labor)
  - Period-to-period access to labor
- Rationale for "self-inflicted pain"

Time-preference rates

- Asymmetric information vs. Full information

  "similar to"

  Positive discounting vs. No discounting

  - Consumption shifts to first period
  - Labor shifts to second period
  - Increase in consumption greater for high valuations
  - Decrease in labor greater for low valuations

- But there are also differences: distribution $F(\theta_t)$ from which valuations are drawn affects consumption

Qualitative conclusions

- "Micro-microfoundations" for intertemporal discounting
- Testable differences: given current valuation, consumption is smaller if individual usually likes the good a lot

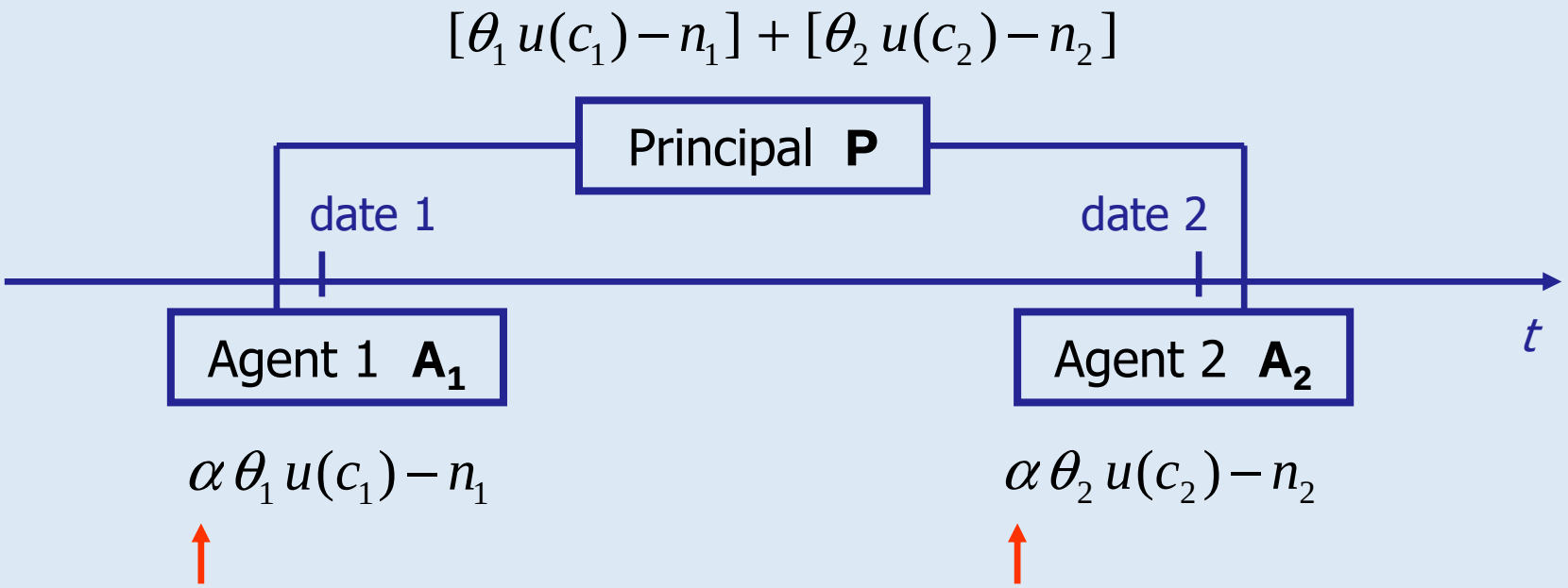
“Incentive salience” and “visceral factors”

Neuroscience: incentive salience

- One system mediates motivation to seek pleasure (wanting)
- A different system mediates the feeling of pleasure (liking)

Stimulus of 1st system ⇨ more work for same reward

Social Psychology: visceral factors (related effect)



$\alpha > 1$: $\mathbf{A}_t$ is tempted to overconsume (biased motivation)

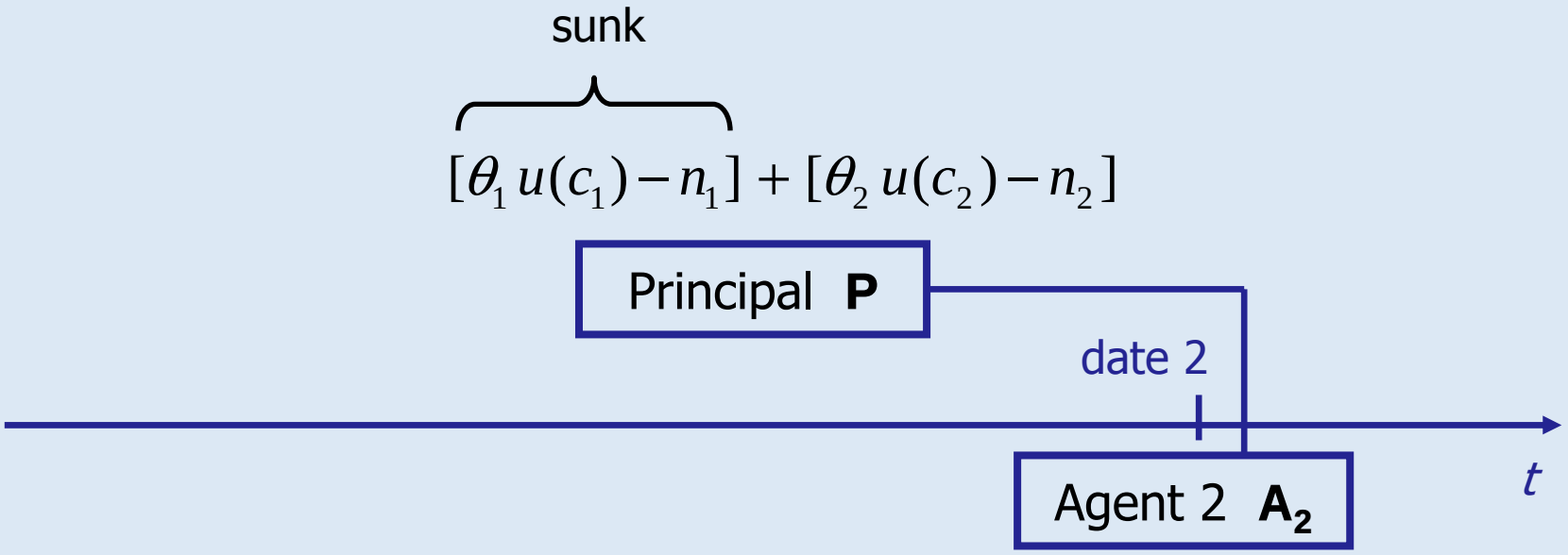
P does not integrate $\mathbf{A}_t$'s salience in "welfare"

Optimal rule at date 1: **P** vs. A_1

very similar to previous case

- **P** offers a menu of consumption and labor pairs
- higher θ_1 ⇨ higher c_1 and higher n_1

Optimal rule at date 2: P vs. A_2



- **P** imposes only two constraints:
  - Consumption cap
  - Budget balance

$$\alpha\theta_2 u(c_2) - n_2$$

- A_2 chooses:
  - If $\theta_2 < \theta^*$: unconstrained optimal pair given his bias
  - If $\theta_2 > \theta^*$: same pair as an agent with valuation θ^*

Qualitative conclusions:

- Optimality requires a simple, non-intrusive rule-of-thumb: "do what you want as long as you don't abuse"
- Stronger bias $(\alpha \uparrow)$ ⇨ tighter control $(\theta^* \downarrow)$

Note: not aware of similar result in any mechanism design problem where **P** has two tools

What's next?

1. Test of behavioral implications
   - Period-to-period labor opportunities affect consumption
   - How much the good is usually liked affects consumption

2. More realistic and comprehensive models of the brain.

We need many more neuro-economic experiments to guide theoretical models:

- Is the hierarchy of the brain "vertical"?
- Is the superior information possessed by the myopic system?
- Are systems with restricted access to knowledge aware of their informational deficit?
- Does the forward looking system discount the future?
- When are salient incentives more likely to operate?